Attorney Docket No. MAILP007

APPLICATION FOR UNITED STATES PATENT

MESSAGE CLASSIFICATION USING ALLOWED ITEMS

By Inventors:

Jonathan J. Oliver
3250 Ash Street
Palo Alto, CA 94306
A citizen of Australia

David A. Koblas
3250 Ash Street
Palo Alto, CA 94306
A citizen of the United States

Brian K. Wilson
3250 Ash Street
Palo Alto, CA 94306
A citizen of the United States

Assignee: MailFrontier, Inc.

VAN PELT AND YI, LLP
10050 N. Foothill Blvd., Suite 200
Cupertino, CA 95014
Telephone (408) 973-2585

MESSAGE CLASSIFICATION USING ALLOWED ITEMS

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. Provisional Patent Application No. 60/476,419 (Attorney Docket No. MAILP007+) entitled A METHOD FOR CLASSIFYING EMAIL USING WHITE CONTENT THUMBPRINTS filed June 6, 2003 which is incorporated herein by reference for all purposes.

This application is a continuation in part of co-pending U.S. Patent Application No. 10/371,987 (Attorney Docket No. MAILP001) entitled USING DISTINGUISHING PROPERTIES TO CLASSIFY MESSAGES filed February 20, 2003, which is incorporated herein by reference for all purposes.

FIELD OF THE INVENTION

The present invention relates generally to message classification. More specifically, a technique for avoiding junk messages (spam) is disclosed.

BACKGROUND OF THE INVENTION

Electronic messages have become an indispensable part of modern communication. Electronic messages such as email or instant messages are popular because they are fast, easy, and have essentially no incremental cost. Unfortunately, these advantages of electronic messages are also exploited by marketers who regularly

send out unsolicited junk messages. The junk messages are referred to as "spam", and spam senders are referred to as "spammers". Spam messages are a nuisance for users. They clog people's inbox, waste system resources, often promote distasteful subjects, and sometimes sponsor outright scams.

There are a number of commonly used techniques for classifying messages and identifying spam. For example, blacklists are sometimes used for tracking known spammers. The sender address of an incoming message is compared to the addresses in the blacklist. A match indicates that the message is spam and prevents the message from being delivered. Other techniques such as rule matching and content filtering analyze the message and determine the classification of the message according to the analysis. Some systems have multiple categories for message classification. For example, a system may classify a message as one of the following categories: spam, likely to be spam, likely to be good email, and good email, where only good email messages are allowed through and the rest are either further processed or discarded.

Spam-blocking systems sometimes misidentify non-spam messages. For example, a system that performs content filtering may be configured to identify any messages that include certain word patterns, such as "savings on airline tickets" as spam. However, an electronic ticket confirmation message that happens to include such word patterns may be misidentified as spam or possibly spam. Misidentification of good messages is undesirable, since it wastes system resources, and in the worst case scenario, causes good messages to be classified as spam and lost.

It would be useful to have a technique that would more accurately identify non-spam messages. Such a technique would not be effective if spammers could easily alter parts of the spam messages they sent so that the messages would be identified as non-spam. Thus, it would also be desirable if non-spam messages identified by such a technique is not easily spoofed.

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention will be readily understood by the following detailed description in conjunction with the accompanying drawings, wherein like reference numerals designate like structural elements, and in which:

Figure 1 is a flowchart illustrating the message classification process according to one embodiment.

Figure 2 is a flowchart illustrating the details of the signature generation process according to one embodiment.

Figure 3 is a flow chart illustrating the classification of a message according to another embodiment.

Figure 4 is a flow chart illustrating a registration process for updating the database, according to one embodiment.

Figure 5 is a table used for aggregating user inputs, according to one system embodiment.

DETAILED DESCRIPTION

It should be appreciated that the present invention can be implemented in numerous ways, including as a process, an apparatus, a system, or a computer readable medium such as a computer readable storage medium or a computer network wherein

program instructions are sent over optical or electronic communication links. It should be noted that the order of the steps of disclosed processes may be altered within the scope of the invention.

A detailed description of one or more preferred embodiments of the invention is provided below along with accompanying figures that illustrate by way of example the principles of the invention. While the invention is described in connection with such embodiments, it should be understood that the invention is not limited to any embodiment. On the contrary, the scope of the invention is limited only by the appended claims and the invention encompasses numerous alternatives, modifications and equivalents. For the purpose of example, numerous specific details are set forth in the following description in order to provide a thorough understanding of the present invention. The present invention may be practiced according to the claims without some or all of these specific details. For the purpose of clarity, technical material that is known in the technical fields related to the invention has not been described in detail so that the present invention is not unnecessarily obscured.

In United States Patent Application No. 10/371,987 (Attorney Docket No. MAILP001) by Wilson, et al filed February 20, 2003 entitled: " USING DISTINGUISHING PROPERTIES TO CLASSIFY MESSAGES" which is herein incorporated by reference for all purposes, a technique using distinguishing properties to identify electronic messages is described. The technique uses distinguishing properties within messages, such as contact information, to identify messages that have previously been classified. In some embodiments, the technique is applied to identify spam

messages. However, spammers aware of such a detection scheme may change their contact information frequently to prevent their messages from being identified.

An improved technique is disclosed. The technique prevents spammers from circumventing detection by using items in the message to identify non-spam messages. All items of a certain type in the message are identified, and checked to determine whether they meet a certain criterion. In some embodiments, the items are distinguishing properties or signatures of distinguishing properties. They are identified and looked up in a database. In various embodiments, the database may be updated by a registration process, based on user input, and/or post-processing stored messages. In some embodiments, the items are looked up in a database of acceptable items. A message is classified as non-spam if all the items are found in the database. If not all the items are found in the database, the message is further processed to determine its classification.

Spammers generally have some motives for sending spam messages. Although spam messages come in all kinds of forms and contain different types of information, nearly all of them contain some distinguishing properties for helping the senders fulfill their goals. For example, in order for the spammer to ever make money from a recipient, there must be some way for the recipient to contact the spammer. Thus, most spam messages include at least one contact point, whether in the form of a phone number, an address, a universal resource locator (URL), or any other appropriate information for establishing contact with some entity. These distinguishing properties, such as contact points, instructions for performing certain tasks, distinctive terms such as stock ticker

symbols, names of products or company, or any other information essential for the message, are extracted and used to identify messages.

Similarly, non-spam messages may also have distinguishing properties. For example, electronic ticket confirmations and online purchase orders commonly include contact points such as URL's, email addresses, and telephone numbers to the sender's organization. It is advantageous that spam messages always include some distinguishing properties that are different from the distinguishing properties in non-spam messages. For example, the URL to the spammer's website is unlikely to appear in any non-spam message. To identify non-spam messages, a database is used for storing acceptable distinguishing properties. The database may be a table, a list, or any other appropriate combination of storage software and hardware. A message that only has acceptable distinguishing properties is unlikely to be spam. Since information that is not distinguishing is discarded during the classification process, it is more difficult for the spammers to alter their message generation scheme to evade detection.

For the purpose of example, details of email message processing using contact points and contact point signatures to determine whether the message is acceptable are discussed, although it should be noted that the technique are also applicable to the classification of other forms of electronic messages using other types of items. It should also be noted that different types of criterion and classification may be used in various embodiments.

Figure 1 is a flowchart illustrating the message classification process according to one embodiment. A message is received (100), and all the contact points are selected (102). It is then determined whether all the contact points can be found in a database of previously stored acceptable contact points (104). If all the contact points are found in the database, the message is classified as non-spam and delivered to the user (106). The contact points that are not found in the database may be contact points for a spammer or contact points for a legitimate sender that have not yet been stored in the database. Thus, if not all contact points are found in the database, the message cannot be classified as non-spam and further processing is needed to accurately classify the message (108). The processing may include any appropriate message classification techniques, such as performing a whitelist test on the sender's address, using summary information or rules to determine whether the content of the message is acceptable, etc.

In some embodiments, the system optionally generates signatures based on the selected contact points. The signatures can be generated using a variety of methods, including compression, expansion, checksum, hash functions, etc. The signatures are looked up in a database of acceptable signatures. If all the signatures are found in the database, the message is classified as non-spam; otherwise, the message is further processed to determine its classification. Since signatures obfuscate the actual contact point information, using signatures provides better privacy protection for the intended recipient of the message, especially when the classification component resides on a different device than the recipient's.

Figure 2 is a flowchart illustrating the details of the signature generation process according to one embodiment. Various contact points are extracted from the message and used to generate the signatures. This process is used both in classifying incoming messages and in updating the database with signatures that are known to be from non-spam. The sender address, email addresses, links to URLs such as web pages, images, etc. and the phone numbers in the message are extracted (200, 202, 204, 206). There are many ways to extract the contact information. For example, telephone numbers usually include 7-10 digits, sometimes separated by dashes and parenthesis. To extract telephone numbers, the text of the message is scanned, and patterns that match various telephone number formats are extracted. Any other appropriate contact information is also extracted (208).

The extracted contact points are then reduced to their canonical equivalents (210). The canonical equivalent of a piece of information is an identifier used to represent the same information, regardless of its format. For example, a telephone number may be represented as 1-800-555-5555 or 1(800)555-5555, but both are reduced to the same canonical equivalent of 18005555555. In some embodiments, the canonical equivalent of an URL and an email address is the domain name. For example, http://www.mailfrontier.com/contact, www.mailfrontier.com/support and jon@mailfrontier.com are all reduced to the same canonical equivalent of mailfrontier.com. It should be noted that there are numerous techniques for arriving at the canonical equivalent of any distinguishing property, and different implementation may employ different techniques.

After the contact points are reduced to their canonical equivalents, signatures corresponding to the canonical equivalents are generated and added to the database (212). There are various techniques for generating the signature, such as performing a hash function or a checksum function on the characters in the canonical equivalent.

The database shown in this embodiment stores signatures that correspond to various acceptable contact points. Such a database is also used in the subsequent embodiments for the purposes of illustration. It should be noted that the acceptable contact points, other distinguishing property and/or their signatures may be stored in the database in some embodiments.

Figure 3 is a flow chart illustrating the classification of a message according to another embodiment. In this embodiment, each contact point of the message is tested and used to classify the message. Once the message is received (300), it is optionally determined whether the message includes any contact points (301). If the message does not include any contact points, the message may or may not be spam. Therefore, control is transferred to 312 to further process the message to classify it. If the message includes at least one contact point, the message is parsed and an attempt is made to extract the next contact point in the message (302). There may not be another contact point to be extracted from the message if all the distinguishing properties in the message have been processed already. Hence, in the next step, it is determined whether the next contact point is available (304). If there are no more distinguishing properties available, the test has concluded without finding any contact point in the message that does not already exist in the database. Therefore, the message is classified as acceptable (306).

If the next contact point is available, it is reduced to its canonical equivalent (307) and a signature is generated based on the canonical equivalent (308). It is then determined whether the signature exists in the database (310). If the signature does not exist in the database, there is a possibility that the message is spam and further processing is needed to classify the message (312). If, however, a signature exists in the database, it indicates that the contact point is acceptable and control is transferred to step 302 where the next contact point in the message is extracted and the process of generating and comparing the signature is repeated.

For the message classification technique to be effective, the database should include as many signatures of acceptable contact points as possible, and exclude any signatures of contact points that may be distinguishing for spam messages. In some embodiments, the database is updated using a registration process. The registration process allows legitimate businesses or organizations to store contact points used in the messages they send to their customers or target audience at a central spam filtering location. The legitimacy of the organization is established using certificates such as the certificate issued by a certificate authority such as Verisign, an identifier or code issued by a central spam filtering authority, or any other appropriate certification mechanism that identifies the validity of an organization.

Figure 4 is a flow chart illustrating a registration process for updating the database, according to one embodiment. Once a registration message is received (400), it is determined whether the certificate is valid (402). If the certificate is not valid, the message is ignored (404). In this embodiment, if the message certificate is valid, optional

steps 405, 406 and 407 are performed. The classification of the message sender is obtained from the certificate (405). It is then further tested using other spam determination techniques to determine whether the message is spam (406). This optional step is used to prevent spammers from obtaining a valid certificate and add their spam messages to the database. If the message is determined to be spam by these additional tests, control is transferred to step 404 and the message is ignored. If, however, the message is determined to be non-spam, one or more signatures are generated based on the contact points in the message (408). The signatures, sender classification, and other associated information for the message are then saved in the database (410).

Different organizations or individuals may have different criteria for which messages are acceptable, and may only allow a subset of the registered signature database. In some embodiments, the signature database from the registration site is duplicated by individual organizations that wish to use the signature database for spam blocking purposes. The system administrators or end users are then able to customize their message filtering policies using the database entries. Using a policy allows some database entries to be selected for filtering purposes.

In some embodiments, the database is updated dynamically as messages are received, based on classifications made by the recipients. Preferably, the system allows for collaborative spam filtering where the response from other recipients in the system is incorporated into the message classification process. Different recipients of the same message may classify the message, therefore the contact points in the message, differently. The same contact point may appear in a message that is classified as non-

spam as well as a message that is classified as spam. The system aggregates the classification information of a contact point, and determines whether it should be added to the database of acceptable contact points.

Figure 5 is a table used for aggregating user inputs, according to one system embodiment. The system extracts the contact points in the messages and generates their signature. The state of each signature is tracked by three counters: acceptable, unacceptable, and unclassified, which are incremented whenever a message that includes the contact point is classified as non-spam, spam or unknown, respectively. A probability of being acceptable is computed by the system based on the counter values and updated periodically. A signature is added to the database once its probability of being acceptable exceeds a certain threshold. In some embodiments, the signature is removed from the database if its probability of being acceptable falls below the threshold.

In some embodiments, the database is updated by post-processing previously stored messages. The messages are classified as spam or non-spam using spam classification techniques and/or previous user inputs. The contact points are extracted, their signatures generated, and their probabilities of being acceptable are computed. The signatures of the contact points that are likely to be acceptable are stored in the database.

An improved technique for classifying electronic messages has been disclosed. The technique uses distinguishing properties in a message and its corresponding signature to classify the message and determine whether it is acceptable. In some embodiments, the distinguishing properties are contact points. A database of registered signatures is